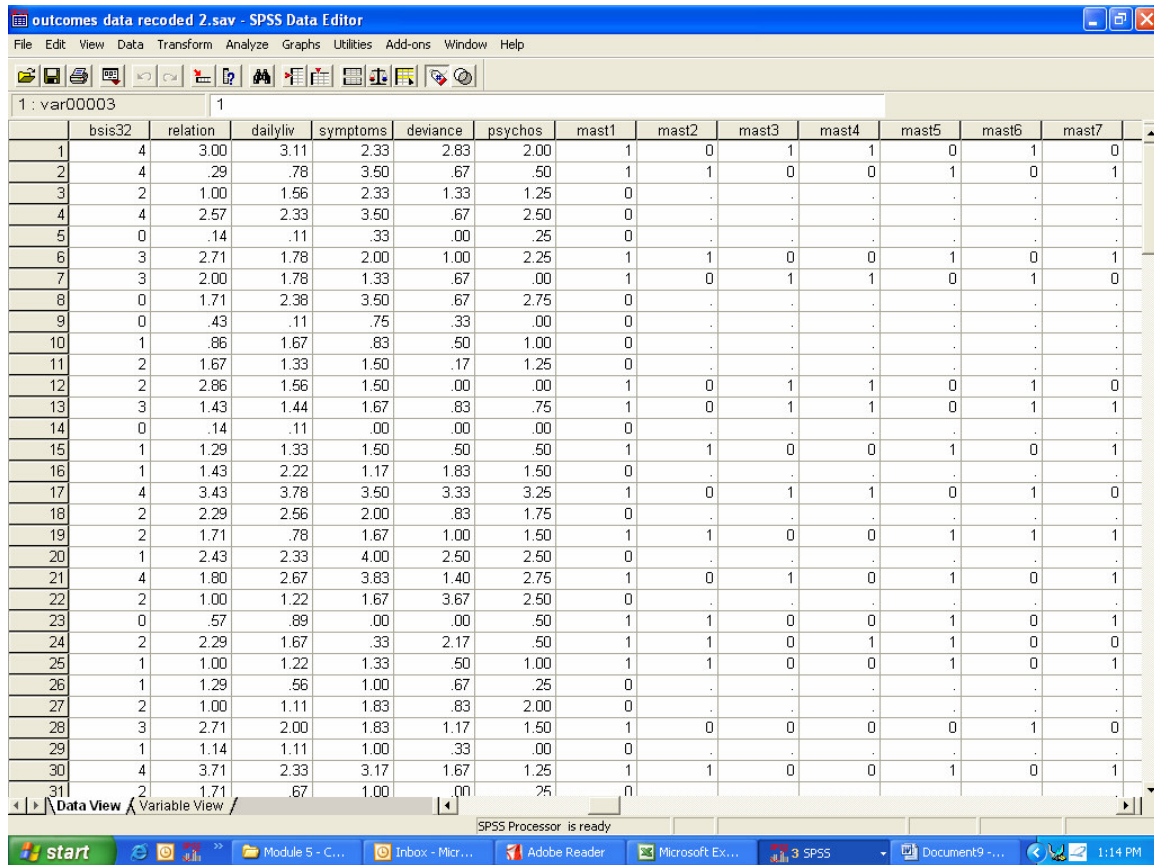


Canonical Correlation in SPSS

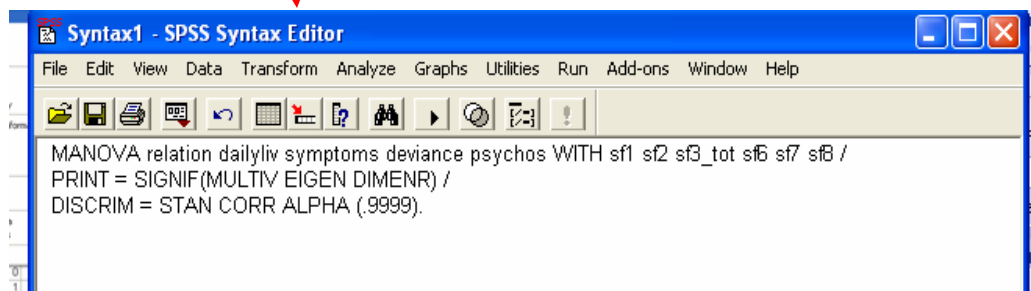
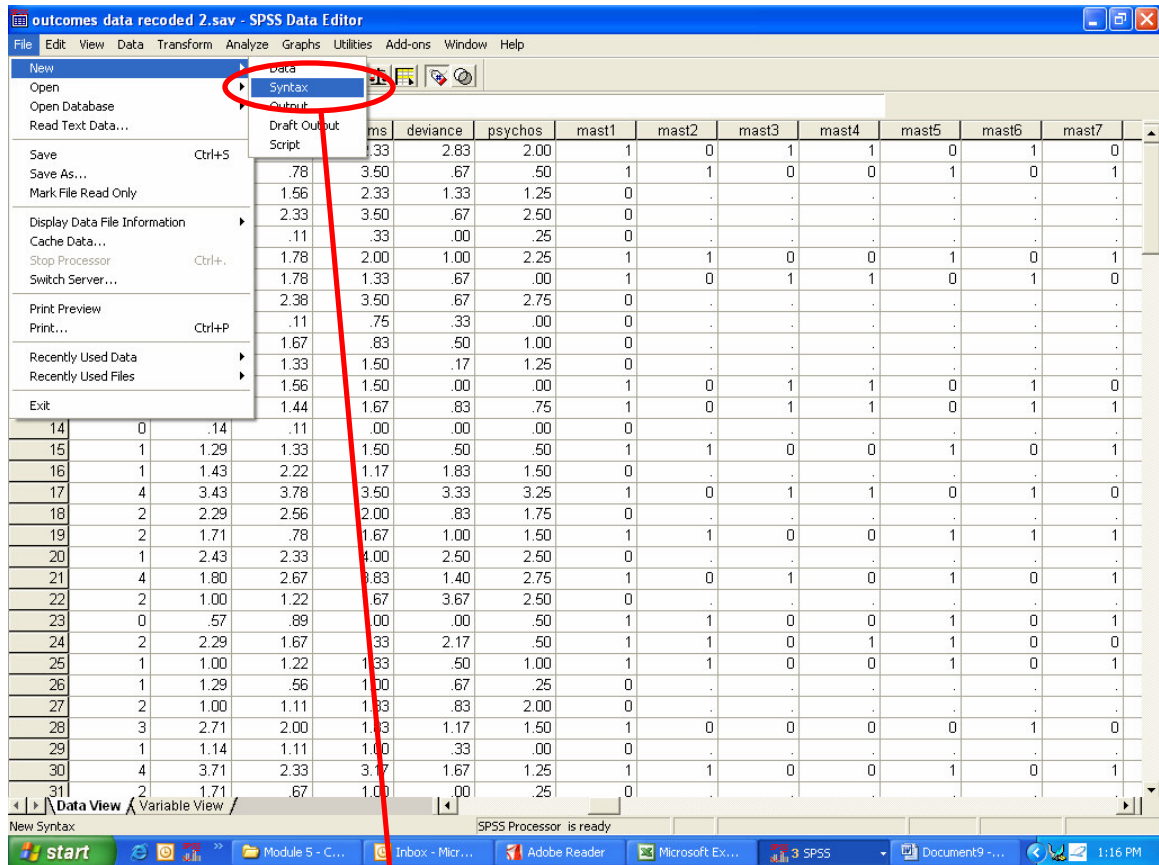
Although these tutorials usually show you how to compute statistics using the menu-driven user interface in SPSS for Windows, I haven't been able to find a menu choice that does canonical correlation. Therefore, we'll have to use SPSS syntax for this one.



	bsis32	relation	dailyliv	symptoms	deviance	psychos	mast1	mast2	mast3	mast4	mast5	mast6	mast7	
1	4	3.00	3.11	2.33	2.83	2.00	1	0	1	1	0	1	0	
2	4	.29	.78	3.50	.67	.50	1	1	0	0	1	0	1	
3	2	1.00	1.56	2.33	1.33	1.25	0	
4	4	2.57	2.33	3.50	.67	2.50	0	
5	0	.14	.11	.33	.00	.25	0	
6	3	2.71	1.78	2.00	1.00	2.25	1	1	0	0	1	0	1	
7	3	2.00	1.78	1.33	.67	.00	1	0	1	1	0	1	0	
8	0	1.71	2.38	3.50	.67	2.75	0	
9	0	.43	.11	.75	.33	.00	0	
10	1	.86	1.67	.83	.50	1.00	0	
11	2	1.67	1.33	1.50	.17	1.25	0	
12	2	2.86	1.56	1.50	.00	.00	1	0	1	1	0	1	0	
13	3	1.43	1.44	1.67	.83	.75	1	0	1	1	0	1	1	
14	0	.14	.11	.00	.00	.00	0	
15	1	1.29	1.33	1.50	.50	.50	1	1	0	0	1	0	1	
16	1	1.43	2.22	1.17	1.83	1.50	0	
17	4	3.43	3.78	3.50	3.33	3.25	1	0	1	1	0	1	0	
18	2	2.29	2.56	2.00	.83	1.75	0	
19	2	1.71	.78	1.67	1.00	1.50	1	1	0	0	1	1	1	
20	1	2.43	2.33	4.00	2.50	2.50	0	
21	4	1.80	2.67	3.83	1.40	2.75	1	0	1	0	1	0	1	
22	2	1.00	1.22	1.67	3.67	2.50	0	
23	0	.57	.89	.00	.00	.50	1	1	0	0	1	0	1	
24	2	2.29	1.67	.33	2.17	.50	1	1	0	1	1	0	0	
25	1	1.00	1.22	1.33	.50	1.00	1	1	0	0	1	0	1	
26	1	1.29	.56	1.00	.67	.25	0	
27	2	1.00	1.11	1.83	.83	2.00	0	
28	3	2.71	2.00	1.83	1.17	1.50	1	0	0	0	0	1	0	
29	1	1.14	1.11	1.00	.33	.00	0	
30	4	3.71	2.33	3.17	1.67	1.25	1	1	0	0	1	0	1	
31	2	1.71	.67	1.00	.00	.25	0	

This dataset shows a large number of variables from a study of homeless individuals. Each participant in this research provided information about their current mental health, physical health, substance use, housing status, and other relevant variables. For this example, let's compute a canonical correlation between some *mental health* variables (subscales of the BASIS-32 measure) and some *physical health* variables (subscales of the SF-36 measure), to see how strong the mind-body relationship is in this dataset.

To create a syntax file in SPSS, go to the “File” menu, and choose “New,” then “Syntax.” The SPSS Syntax Editor will open in a separate window.



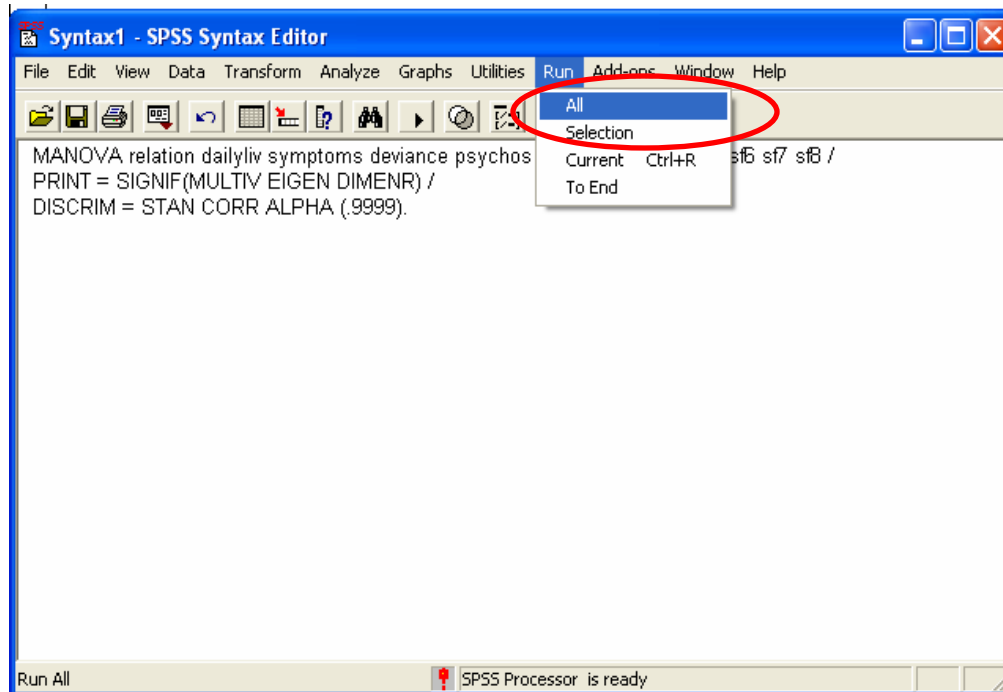
The predictor variables listed in this syntax are subscales of the BASIS-32 showing quality of interpersonal relationships (relation), activities of daily living (dailyliv), anxiety/depression symptom severity (symptoms), antisocial behavior (deviance), and psychotic symptoms (psychos). The criterion variables are six different health items or subscales from the SF-36 (sf1, sf2, sf3_tot, sf6, sf7, and sf8).

Type the following in the Syntax Editor window (substitute the names of the variables you are interested in for the ones I have here in lower case type – the idea is to put in a list of your IVs with a space between each of them, then the word WITH, then the list of your DVs). Make sure that you have a regular slash (/) at the end of each line – this tells SPSS that whatever comes next is a subcommand. Also make sure that you have a period (.) at the end of the last command line – this tells SPSS that it has reached the end of the command sequence. Here is the syntax to use:

```
MANOVA relation dailyliv symptoms deviance psychos WITH sf1 sf2 sf3_tot sf6 sf7 sf8 /  
PRINT = SIGNIF(MULTIV EIGEN) /  
DISCRIM = STAN.
```

Again, the lower-case entries on the first line are just the variable names, so if you are doing this procedure with something other than my sample dataset, you should put in the names of your IVs and DVs, exactly the way those variables are named in the SPSS data file you are working with. The rest of the syntax consists of SPSS commands, telling the program what type of output to give you, and will be the same whenever you run this procedure. You can even cut and paste from this document into the Syntax Editor, and then just change the variable names as needed.

Once you have completed the syntax, go to the “Run” menu (still in the Syntax Editor – this menu doesn’t appear in the main SPSS window), and select “All” to tell SPSS to run your entire sequence of syntax commands.



SPSS will automatically open an output file and show you the results. If for some reason you get an error message instead of a result, go back and double-check your syntax, especially the slash at the end of the first two lines and the period at the end of the third.

Here are selected results of the procedure:

```
* * * * * A n a l y s i s   o f   V a r i a n c e  -- design  1 * * *
* * *
```

```
EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 5, M = 0, N = 56 )
```

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.56569	2.50890	30.00	590.00	.000
Hotellings	.83412	3.12516	30.00	562.00	.000
Wilks	.50739	2.82210	30.00	458.00	.000
Roys	.39437				

This first section contains some multivariate tests for statistical significance – skip over them for now. These are explained in more detail in the MANOVA procedure.

```
Eigenvalues and Canonical Correlations
```

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	.651	78.066	78.066	.628	.394
2	.092	11.001	89.067	.290	.084
3	.050	6.042	95.109	.219	.048
4	.039	4.733	99.841	.195	.038
5	.001	.159	100.000	.036	.001

This is the actual listing of canonical correlations. There is more than one way to combine the IVs into a latent factor, to combine the DVs into a latent factor, and to relate the two latent factors to one another. The first canonical correlation is always the largest, because it was selected to maximize the associations between the two sets of variables. That's the one that is usually reported for a canonical correlation analysis. The right-hand column gives the “squared correlation” (r^2), which is the % of variability in all the DVs together that can be accounted for by all the IVs together.

```
EFFECT .. WITHIN CELLS Regression (Cont.)
Univariate F-tests with (6,118) D. F.
```

Variable	Sq. Mul. R	Adj. R-sq.	Hypoth. MS	Error MS	F
relation	.27579	.23897	5.32880	.71150	7.48951
dailyliv	.37424	.34242	6.57706	.55919	11.76180
symptoms	.27276	.23578	7.03305	.95350	7.37605
deviance	.13368	.08963	2.55777	.84281	3.03480
psychos	.20741	.16711	4.71506	.91614	5.14664

Variable	Sig. of F
relation	.000
dailyliv	.000
symptoms	.000
deviance	.009
psychos	.000

This third section shows you an *F*-test for each individual dependent variable, showing whether it is significantly related to the latent IV.

```
* * * * * A n a l y s i s   o f   V a r i a n c e  -- design   1 * * *
* * *
```

```
Standardized canonical coefficients for DEPENDENT variables
      Function No.
```

```
Variable           1

relation           .054
dailyliv           .790
symptoms           .156
deviance           -.289
psychos            .305
```

This table (produced by the “DISCRIM = STAN” subcommand) shows you the standardized canonical loadings for each of the dependent variables. The variables with the highest canonical loadings are the most strongly related to the latent DV – so in this case, “dailyliv” has the strongest association with the underlying construct, and “relation” the lowest association with it. “Deviance” is negatively correlated – the more deviance, the lower the score on the latent DV.

```
Standardized canonical coefficients for COVARIATES
      CAN. VAR.
```

```
COVARIATE           1

sf1                -.164
sf2                -.302
sf3_tot            -.031
sf6                -.668
sf7                .372
sf8                -.556
```

This is the same information for the IVs. In this case, all of the questions except sf7 are negatively related to the underlying construct – if the underlying construct is “health,” this means that higher scores on the questions mean lower levels of health (which is actually how the SF-36 questions are set up). Variables sf6 and sf8 have the strongest relationships (both negative, but strong) with the underlying construct of “health.”